Cross-modal perception of puppies and adult conspecifics in dogs (*Canis familiaris*)

Yuri Kawaguchi [1, 2a], ORCID: 0000-0002-6682-4964

Zsófia Virányi [1],

Tamás Faragó [3], ORCID: 0000-0001-5987-2629

Ludwig Huber [1], ORCID: 0000-0002-0217-136X

Christoph J. Völter [1], ORCID: 0000-0002-8368-7201

1 Comparative Cognition, Messerli Research Institute, University of Veterinary Medicine Vienna, Medical University of Vienna and University of Vienna, Vienna, Austria

2 Japan Society for the Promotion of Science, Tokyo, Japan

3 Neuroethology of Communication Lab, Department of Ethology, Eötvös Loránd University, Budapest, Hungary

[a] Correspondence concerning this article should be addressed to Yuri Kawaguchi, who is currently a Newton International Fellow at Evolution and Social Interaction Research Group, School of Social Sciences, Nottingham Trent University.
(Email: yuri.kawaguchi.09@gmail.com, Address: School of Social Sciences, Nottingham Trent University, 50 Shakespeare Street Nottingham NG1 4FQ, Nottingham, UK)

Acknowledgement

We would like to thank all the dogs and dog caregivers who participated, Laura Laussegger and Marion Umek, for helping with eye-tracking data collection, and the members of Clever Dog lab, especially Karin Bayer, for their support and comments on the study. We also thank Hoi-Lam Jim for her English help. This project was supported by Japan Society for the Promotion of Science (JSPS) for Y. K. T. F. was supported by the Hungarian Academy of Sciences via the János Bolyai Research Scholarship (BO/751/20), the ÚNKP-22-5 New National Excellence Program of the Ministry for Innovation and Technology from the source of the National Research, Development and Innovation Fund (ÚNKP-22-5-ELTE-475) and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (950159). We also thank the editor and the two anonymous reviewers for their very constructive comments and suggestions.

Abstract

Understanding conspecifics' age classes is crucial for animals, facilitating adaptive behavioral responses to their social environment. This may include gathering and integrating information through multiple modalities. Using a cross-modal preferential looking paradigm, we investigated whether dogs possess a cross-modal mental representation of conspecific age classes. In Experiment 1, dogs were presented with images of an adult dog and a puppy projected side-by-side on a wall while a vocalization of either an adult dog or a puppy was played back simultaneously. To test the effect of relative body size between adult dog and puppy images, two size conditions (natural size and same size) were employed for visual stimuli. We examined dogs' looking behavior in response to cross-modally matched versus mismatched stimuli. We predicted that if dogs have cross-modal representations of age classes, they would exhibit prolonged attention towards matched images compared to mismatched ones. In Experiment 2, we administered the same paradigm within an eye-tracking experiment to further improve the measurement quality of dogs' looking times. However, dogs' looking times in either experiment did not demonstrate significant differences based on the match or mismatch between image and vocalization. Instead, we observed a size effect, indicating dogs' increased attention towards larger adult dog images compared to smaller puppy images. Consequently, we found no evidence of cross-modal representation of age class in dogs. Nonetheless, we found increased looking time and pupil size upon hearing puppy vocalizations compared to adult vocalizations in Experiment 2, suggesting that dogs exhibited heightened arousal when hearing puppy whining.

Keywords: cross-modal recognition, dogs, puppy, eye tracking, whining

Acquiring information about conspecifics is important, especially for social animals. Animals achieve this by using multiple sensory modalities; for example, non-human animals integrate information about species (Adachi et al., 2009; Gergely et al., 2019; Mongillo et al., 2021), identity (conspecific: Proops et al., 2009, human: Adachi et al., 2007; Takagi et al., 2019), human gender (Ratcliffe et al., 2014), and conspecific body size (Bálint et al., 2013; Faragó, Pongrácz, Miklósi, et al., 2010; Taylor et al., 2011). Recognizing if a conspecific is an adult or an infant is also important for social animals, as it enables them to adapt their behavior to the individuals around them (Berry & McArthur, 1986). Animals should not treat an adult individual like an infant and vice versa. For example, animals should avoid immature individuals as mating partners, and unfamiliar individuals can be a threat if they are adults, but probably not if they are infants. Additionally, approaching others' young infants often provokes aggressive behavior from the parents. Thus, ignoring information about others' age can be very costly.

Which cues do animals use to recognize others' age class? Information exchange about this requires two things: 1) information about age class should be expressed in certain ways, and 2) these cues (or signals) should be appropriately perceived and used by other individuals. Age cues can be visual (e.g., body size, pelage or skin color, morphology, or behavior), auditory (e.g., acoustic characteristic of vocalizations), or olfactory (e.g., body smell). Previous studies have shown that animals can use some of these cues; for example, Ghazanfar et al. (2007) tested whether rhesus macaques (*Macaca mulatta*) could match adult versus juvenile visual and acoustic information by using the cross-modal matching paradigm. The monkeys saw videos of adult and juvenile conspecific heads while they heard matching or non-matching vocalizations. The results provided evidence that monkeys can integrate information about vocalizations and images of juvenile and adult conspecifics, different in body size. Age class information is also important for Australian sea lions (*Neophoca cinerea*). For instance, females can be very aggressive toward

non-filial pups if they approach; therefore, approaching a non-mother adult female can be risky for pups. Charrier et al. (2022) found that when replicas of pups and adult females were presented to Australian sea lions, pups avoided female replicas but not pup replicas. Female adults also showed a clear preference toward the pup replica matching their own pup's age class compared to one of a different age class, indicating that they can use visual cues to differentiate conspecific age class.

Dogs (*Canis familiaris*) are an especially interesting species in which to test cross-modal age recognition for conspecifics because their appearance varies greatly among breeds, including body size, which is a potential cue of age class in many species. Adult dog weight varies from less than one kg to 100 kg among breeds (Autier-Dérian et al., 2013), comparable to or even bigger than the developmental changes each dog experiences during its lifetime. Consequently, one possibility is that across different breeds, body size may not consistently serve as a reliable indicator of age class in dogs. On the other hand, it is also possible that the inclination to use body size as a proxy of age class is still preserved in dogs and dogs may rely on body size. Specifically, they might rely on the relative body size between a mother and her pup to differentiate between age classes.

Body size recognition in dogs has been cross-modally tested, and studies have shown that dogs are sensitive to conspecific body size information (Faragó, Pongrácz, Miklósi, et al., 2010; Taylor et al., 2011). One common way to test cross-modal recognition is by using a preferential-looking paradigm, in which a pair of visual stimuli is presented with one type of sound, and then attention to cross-modally matched versus mismatched visual stimuli are compared. If subjects pay more attention to visual stimuli that match the acoustic stimuli, it implies that they have a cognitive representation of the object which links to the two perceptual cues across modalities (e.g., Faragó, Pongrácz, Miklósi, et al., 2010; Sliwa et al., 2011). Faragó

et al. (2010) tested whether dogs represented body size after hearing conspecific growls. They replayed a conspecific growl and simultaneously projected two images of a dog side-by-side. The two images were identical in content but different in size; one was sized to show the growling dog's actual size, while the other was either 30% larger or smaller. They found that dogs looked at the size-matched image sooner and longer upon hearing the growl. Further evidence for cross-modal matching of body size and growls in dogs has been reported by Taylor et al. (2011). These results suggest that dogs have a visual representation of body sizes when they hear conspecific growls; that is, dogs can gain information about conspecific body size from both visual and auditory cues. However, it remains unclear if dogs use body size information as cues of age class, considering the huge variation of body size even among adult dogs of different breeds. Consequently, even if conspecific body size is important information for dogs, they may not rely on it to differentiate puppies and adult dogs but use other cues such as body proportion or vocalization (both vocalization types and acoustic properties).

Dogs' responses to both conspecific and heterospecific infant vocalizations have been tested, and these previous studies have found that dogs are responsive to them. For example, Yong and Ruffman (2014) tested dogs' cortisol levels after hearing either a human infant crying, babbling or white noise. They found that, like in humans, dogs' cortisol levels significantly increased from the baseline after hearing an infant crying, and they interpreted this as evidence of emotional contagion in dogs susceptible to human infant crying. However, dogs are even more responsive to vocalizations of conspecific pups than human infant cries. Lehoczki et al. (2020) played different kinds of sound, including separation calls of dog pups, kittens, human infants, and artificial sounds and tested dogs' responses. Upon hearing the separation calls, dogs oriented their head to the sound source quicker when the sound was a dog pup call or artificial sound than a human infant cry or kitten call. Root-Gutteridge et al. (2021) also reported that dogs were more

attentive to vocalizations in the call-frequency range of conspecific pups than that of human infants. These results indicated that particular acoustic characteristics of pup calls especially attract dogs' attention. Importantly, however, none of these studies compared the responses toward adult versus infant vocalizations. Therefore, even though dogs react to puppy or human infant vocalizations, it remains unclear whether dogs recognize that they were produced by infants but not adults or even whether the response is specific to vocalizations of young animals or distress calls in general.

In this study, we investigated dogs' capacity for cross-modal recognition of puppy and adult conspecifics by using a preferential-looking paradigm. We played back vocalizations of either an adult dog or a puppy while simultaneously presenting images of an adult dog and a puppy. We then coded which image the dogs first looked at and also compared the durations of looking at the two images. We predicted that if dogs have a mental representation of puppies and adult conspecifics that includes both auditory and visual properties, dogs will look at the matched image sooner and/or for longer than the mismatched image. The looking time measure is also used as an index of expectancy violation in some studies testing cross-modal recognition (e.g., Adachi et al., 2007; Takagi et al., 2019). Importantly, the majority of those studies reporting a bias for incongruent stimuli used a single visual stimulus (i.e., either congruent or incongruent) instead of paired visual stimuli like ours (but see also Jardat et al., 2022). Therefore, in the present study with a preferential looking paradigm, we predicted a looking bias for a congruent image. We prepared two size conditions for visual stimuli to differentiate relative body size cues and other cues. In the natural size condition, an adult dog image was presented as bigger than a puppy, while in the same size condition, images of an adult and a puppy were presented in the same size by changing the size of both. If dogs use visual cues other than size, we predicted a congruency effect in both conditions. However, if body size serves as an important visual cue to differentiate between

the adult and puppy, we expected a congruency effect in the natural size condition but no or a diminished effect in the same size condition. A previous cross-modal matching study reported that dogs' bias for congruent stimuli depended on the subjects' previous experience. In Ratcliffe et al. (2014), where either a male or female voice played in the presence of a man and a woman, dogs living with more than two adults showed looking bias for the congruent person while dogs living with one or two persons showed a bias for the incongruent person. Thus, it is possible that only dogs who have had previous experience with both adult dogs and puppies show a congruency effect. In order to take this possible influence into account, we asked the owners about their dog's previous experience with puppies. In Experiment 1, based on the methods of Faragó et al. (2010), the visual stimuli were projected side-by-side on a wall and the looking behavior of the subjects was recorded by video cameras. In Experiment 2, we applied eye-tracking using the same paradigm, in order to test the results in Experiment 1 when analyzing the dogs' gaze behavior in a more detailed manner.

Experiment 1

Methods

Subjects

We tested 43 adult pet dogs; five dogs were excluded because of technical issues with the recording. The data was collected from 23rd September to 12th October 2021. The sample size was determined based on a power simulation. Our power simulation was based on 36 subjects, yielding a power of over 80%. Therefore, we aimed for this sample size. Sessions were excluded if the dog did not look at either image in two or more trials. Based on this pre-determined criterion, four dogs were excluded. Therefore, the final sample size consisted of 34 dogs (17 females, M = 71.6 months, see Table A1 for details). The owners were informed about the study and gave

158 written consent before participating, but the purpose of the study was explained after. We asked
159 the owners about their dogs' previous experience of interacting with neonate puppies (3 answers
160 were possible: not at all: "*My dog has never met a puppy*.", some experience: "*My dog has met a*
161 *puppy several times."*, daily experience: *"My dog has interacted with one or more puppies with*
162 *a daily basis." (e.g. lived with puppies)).*

163 Experimental set-up and Procedure

164 The experiment was conducted in an experimental room (6m x 7m) at the Clever Dog
165 Lab (Messerli Research Institute, University of Veterinary Medicine Vienna, Austria). The owner
166 sat on a chair in the middle of the experimental room and gently held the dog so that her/his head
167 was in the center of the square marked on the floor (Figure 1). Tape was used to mark the floor
168 for visual guidance for later head direction coding. Two projector screens (150 cm x 150 cm each)
169 were placed by the wall the dogs faced and sat from 3 m away. A loudspeaker (Dell AC511
170 Soundbar) was placed centrally between the screens. The visual stimuli were presented on the
171 two screens at floor level by a projector (EPSON LCD Projector) located behind the subject and
172 the owner. Four cameras recorded the dogs' behavior and experimental procedures: one was
173 attached on the ceiling above the subject, one was located between the two screens and the other
174 two were attached to the side walls (Figure 1). The lights were turned off during the experiment,
175 and the window blinds were closed. A visual occluder was placed in front of the doors to minimize
176 distraction from the outside.

177 After the subject was familiarized with the experimental room, we started the
178 experiment. To prevent the owner from influencing their dog's behavior, they were not allowed
179 to see the visual stimuli, so they either closed their eyes or looked down. Each trial started with
180 the presentation of an attention-getter, a short animation projected on the wall between the two
181 screens together while replaying a short sound. When the subject was facing the center of the wall,

the experimenter immediately started projecting two images of an adult dog and a puppy on the two screens side-by-side. At the same time, a recording of a dog vocalization (either adult dog or puppy) was played through the loudspeaker. Each image pair was presented for 10 seconds, and the vocalization for approximately 3 seconds. The previous study by Faragó et al. (2010) presented visual stimuli for 20 seconds, but we shortened the duration to keep the dogs' concentration. Each session consisted of four trials, and each dog was tested in two sessions (same size or natural size sessions, see below) on the same day with a short break between sessions. The study was approved by the institutional ethics and animal welfare committee (ETK-106/07/2021) in accordance with GPS and national legislation guidelines. The experimental design and analyses in Experiment 1 were pre-registered on AsPredicted (https://aspredicted.org/GJ1_X5G).

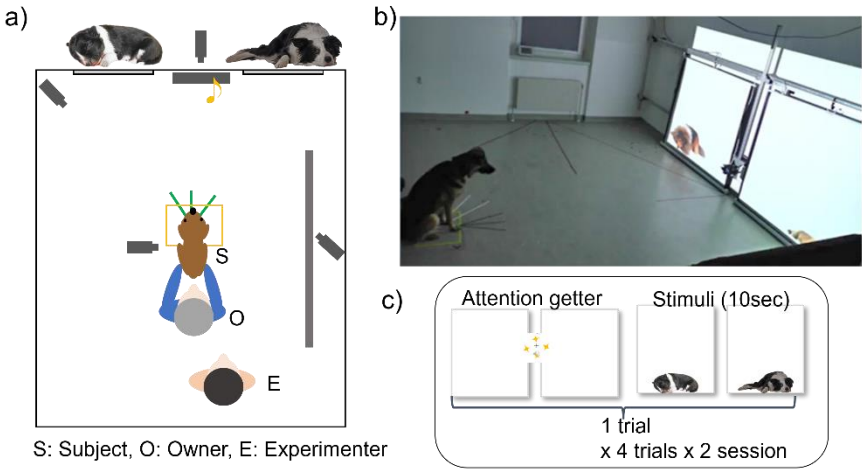

Figure 1 The experimental setting

a) a schematic image of the experimental room, including four cameras, two screens, a loudspeaker, and a visual occluder. b) a dog participating in the experiment. c) an example of one trial.

198 Stimuli

199 **Visual stimuli**

200 We used images of adult dogs (>1 year; n = 8) and puppies (<1 month; n = 8). Since we
201 did not restrict the subjects' breeds, their breeds and the body size varied. In order to avoid the
202 potential risk of using only one breed as stimuli, we used dogs from two breeds as stimuli. Namely,
203 half of the stimuli depicted border collies, representing a medium-sized breed and the other half
204 Shetland sheepdogs, representing a small-sized breed. Sex and the exact age of some dogs in the
205 images were unknown. The eyes of the puppies were closed because they were neonate, while the
206 adult dogs' eyes were open considering the possibility that it may be unnatural if a "sleeping"
207 adult dog vocalized. The dogs presented as the stimuli were in a lying position, which was chosen
208 to have the same posture for both adult dogs and puppies. The brightness was matched between
209 an adult dog and a puppy for each pair, and the background was trimmed. In the natural size
210 condition, the adult image was always larger than the puppy image (i.e., approximately three times
211 larger in height) for both stimuli breeds, while in the same size condition, an adult dog and a
212 puppy were presented in the same size (about 20% smaller in height than the adult dog of natural
213 size condition). The size condition was blocked into sessions, and the order of the sessions (i.e.,
214 same/natural size conditions) was counterbalanced. The stimuli breed type was mixed in each
215 session, but the order was counterbalanced between subjects (e.g., the first two trials were border
216 collie in session 1 and Shetland sheepdog in session 2 for half of the subjects). The side of the
217 adult dog image was balanced so that the matched image side varied across trials.

218

219 **Auditory stimuli**

220 We used whining sounds recorded from four adult dogs and four neonate puppies.
221 Whines are high-pitched tonal vocalization which puppies emit when they are separated from

222 their mother, but they are also produced by adult dogs upon separation from the owner or in
223 different situations (Yeon, 2007). Whines are one of the earliest sounds present in canids,
224 developmentally proceeding the emergence of growls (Cohen & Fox, 1976). Thus, whines were
225 chosen because this is a type of vocalization which both adult dogs and puppies make. Half of
226 them were from Border collies, and the other half were from Shetland sheepdogs, and the breeds
227 of visual and auditory stimuli were always matched in each trial. The stimuli were provided by
228 the pup-sound database of the Department of Ethology（Eötvös Loránd University, Budapest,
229 Hungary); they had been pre-recorded and used for other studies (Lehoczki et al., 2019, 2020;
230 Marx et al., 2021). The adult vocalizations were recorded when the owner left the dog alone in
231 the experimental room for 3 min. The puppy vocalizations were recorded at 4 days age at the
232 breeders' home during a 3 min separation from the mother and littermates while only the
233 experimenter and the breeder were present. The duration of the stimuli used was two to three
234 seconds. The vocalizations were normalized among the stimuli to control loudness variations
235 using the software Audacity (https://www.audacityteam.org/). For acoustic details see
236 Supplementary Material.
237
238 Analysis
239 Dogs' looking behavior was coded using the Loopy scoring tool (loopbio gmbh, Vienna,
240 Austria). The coder (Y. K.) knew the purpose of the study but was blind to the vocalization type,
241 hence the matching image's side for each trial during the coding. Looking was defined as the head
242 oriented to one of the images. Whenever the orientation of the eyes was clearly visible, it was
243 considered. If a dog turned her/his head or gazed continuously (i.e., without stopping the gaze for
244 longer than 0.1 sec) it was not coded as looking at the corresponding stimulus. A second coder
245 who was naïve to the hypothesis and the prediction of the study coded 20% of the valid sessions

(10 sessions) to check inter-observer reliability, which was good (intraclass correlation coefficient ICC (2,1) = 0.83).

We investigated whether dogs' visual attention towards images of adult and puppy dogs was influenced by vocalization types and body size. If dogs displayed greater attention to cross-modally matched images, their looking behavior would be biased towards one image (e.g., an adult image) compared to the other when they heard corresponding (adult) vocalization (indicated by a significant effect of *vocalization*). Additionally, if dogs exhibited increased attention to cross-modally matched images only when the relative body size between an adult dog and a puppy is preserved, we expected their looking behavior to be biased towards the corresponding image (e.g., an adult image) only in the condition where images were presented at the natural size (indicated by a significant interaction effect between *vocalization* and *size*). Following our pre-registered analysis plan, we analyzed the proportion looking time to the adult image compared to the puppy image (*adult looking proportion*) during the entire trial duration. Adult looking proportion was calculated by adult looking time divided by the sum of adult looking time and puppy looking time in each trial. We also recorded which image the dog looked at first (*the first look image*, adult coded as 1; puppy coded as 0). We fitted Generalized Linear Mixed-Effects Models (GLMMs) using R (R Core Team, 2018) and the packages glmmTMB (Brooks et al., 2017) and lme4 (Bates et al., 2015). We used a beta distribution (continuous responses between 0 to 1) for the proportion of looking time and a binomial error distribution for the binary first look variable. The following factors were included as predictor variables: size condition (natural size/same size), vocalization (adult dog/puppy), the interaction between these two terms, stimuli breed type (border collie/Shetland sheepdog), the side of the adult image (left/right), the session number (1 or 2), the trial number within a session (1-4), subject age, and sex. Subject ID was included as a random intercept, and size condition, vocalization, stimuli breed type, side of the adult image, session

number, and trial number were included as random slopes. The model syntax in R was: `dependent variable ~ size condition*vocalization + stimuli breed type + adult side + session + trial + age + sex + (1+ condition + vocalization + stimuli breed type + adult side + session+ trial ||subject id)`. The random-effects structure was kept maximal to maintain conservativity (Barr et al., 2013). We also checked models with a simplified random slope structure (i.e., removing stimuli breed type, adult side, session, trial, age, sex), and the results did not change the pattern of significant results (see Supplementary Material). We conducted likelihood ratio tests (using R function drop1 with argument test set to "chisq") to evaluate the significance of the fixed effects.

In the adult looking proportion analysis, information about absolute looking times was removed (e.g., a dog that looked for 1 s at the adult image or for 10 s had the same value of 1 if they did not look at the puppy image). Indeed, only in 31.3 % of the valid trials, dogs looked at both images. In order to take into account the possibility that vocalization affected absolute looking time to match and mismatch, we conducted further exploratory analyses in addition to the pre-registered analyses. We analyzed adult looking time and puppy looking time divided by the total presentation duration (i.e., 10 s) using a beta GLMM. Moreover, we also tested the effect of dogs' previous experience of interacting with puppies. We analyzed match looking proportion, defined as looking time to the matched image divided by the sum of looking time to matched images plus mismatched images, by fitting the same beta GLMM but with experience as an additional predictor variable. The data, R code and a video are available at Mendeley Data (doi: 10.17632/6skk6w8tnb.1).

Transparency and Openness Statement

- Data availability
  The data, R code and a video are available at Mendeley Data, doi: 10.17632/6skk6w8tnb.1
- Analysis code availability

The data, R code, a video and the results of the acoustic analysis are available at Mendeley Data, doi: 10.17632/6skk6w8tnb.1

- Materials availability
  The stimuli we used are available upon request.
- Citation (to secondary data, materials, and/or code, including statistical packages)
  The data, R code, a video the results of the acoustic analysis are available at Mendeley Data, doi: 10.17632/6skk6w8tnb.1
- Reporting standards
  We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study.
- Design preregistration availability
  The experimental design and analyses in Experiment 1 were pre-registered on AsPredicted (https://aspredicted.org/GJ1_X5G).
- Analysis plan preregistration availability
  The experimental design and analyses in Experiment 1 were pre-registered on AsPredicted (https://aspredicted.org/GJ1_X5G).

Results

To test whether dogs looked at cross-modally matched images more than the mismatched ones, we examined whether dogs' looking behavior (the adult looking proportion and the first look) was influenced by vocalization types with the interaction with size condition. Figure 2 shows the adult looking proportion (see also Figure A1 for absolute looking time). For the analysis of adult looking proportion, the interaction between size condition and vocalization was not significant (beta GLMM, likelihood ratio (LR) test: $\chi^2(1) = 0.87$, $p = 0.35$). This indicates that the effect of vocalization type on the dog's looking times did not differ significantly between the two body size conditions. Thus, we excluded it from further analysis and fitted another model. The results revealed that the vocalization type (adult dog/puppy) did not significantly affect adult looking proportion ($\chi^2(1) = 0.61$, $p = 0.43$, Table A3 and A4). The effects of size condition ($\chi^2(1) = 9.29$, $p = 0.002$) and adult side ($\chi^2(1) = 8.58$, $p = 0.003$) on adult looking proportion were

significant, indicating that the dogs looked longer at the adult image when it was larger than the puppy image and when it was presented on the right side. The GLMM of the first look provided no evidence for a significant interaction between size condition and vocalization (binomial GLMM, LR test: $\chi^2(1) = 0.32$, $p = 0.57$). After removing the interaction, we found a significant effect of size condition ($\chi^2(1) = 3.94$, $p = 0.047$, Figure 2, Table A3 and A4) but not of vocalization ($\chi^2(1) = 1.43$, $p = 0.23$) on the first look. This demonstrates that the likelihood of dogs looking at the adult image first was higher when it was larger than the puppy image (natural size condition) than when they were the same size (same size condition).

In an exploratory analysis that was not pre-registered, we analyzed adult looking time and puppy looking time divided by the total presentation duration; however, the main effect of vocalization was not significant for either response variable (beta GLMM, LR test: adult looking time: $\chi^2(1) = 0.48$, $p = 0.49$, puppy looking time: $\chi^2(1) = 0.001$, $p = 0.97$, Table A5 for details). We conducted a further analysis to investigate if dogs' visual attention towards cross-modally matched images, compared to mismatched ones, is influenced by their previous exposure to puppies. We tested the effect of dogs' previous experience of interacting with puppies on the match looking proportion, which represents the ratio of the looking time directed at matched images to the combined looking time at both matched and mismatched images. Based on the owners' answers, sixteen dogs had no experience with a young puppy, fifteen had some experience, and three had daily experience. Since only a few dogs had daily experience, we combined dogs with some or daily experience of puppies (18 dogs, "experienced dogs") in comparison to dogs with no experience of puppies (16 dogs, "inexperienced dogs"). Nevertheless, the effect of experience on the match looking proportion was not significant (beta GLMM, LR test: $\chi^2(1)$

=0.07, $p = 0.79$, Table A6 for details).

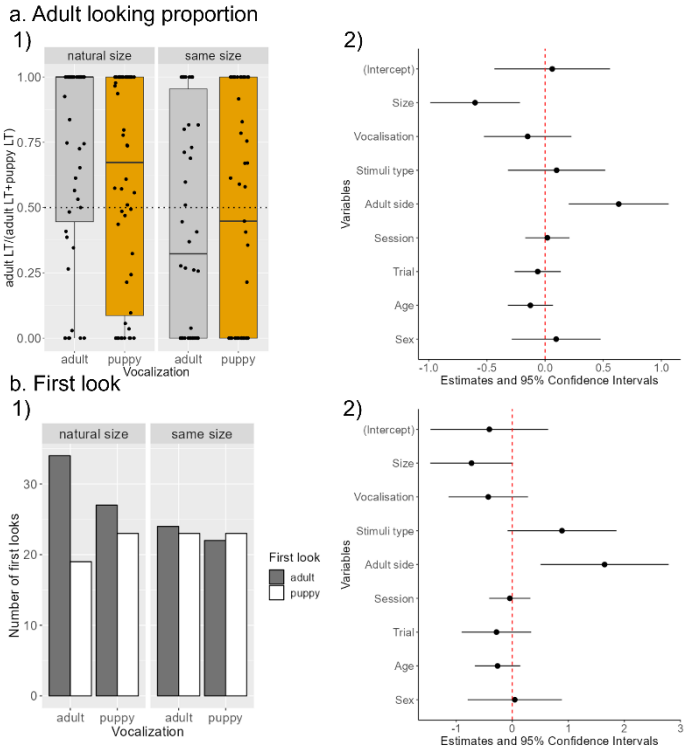

Figure 2 The results of looking behavior in Experiment 1

a-1) Boxplot showing the proportion of looking at the adult dogs (adult looking time/ (adult looking time + puppy looking time). The dots represent the data point for each trial. Scores greater than 0.5 in the adult looking proportion indicate that a subject looked at an adult dog for a longer time than at a puppy. a-2) Estimate of fixed effects and 95% confidence interval of the GLMM modelling the adult looking proportion.

b-1) The total number of first looks from all subjects. b-2) Estimate of fixed effects and 95% confidence interval of the GLMM modelling the first looks.

Discussion

The analysis of looking time proportion or first look did not yield a significant effect of vocalization type (adult dog/puppy), indicating that dogs did not change their looking behavior toward the adult or puppy images based on whether the vocalization was matched or not. Thus, the results provided no evidence of cross-modal representation of age class in dogs. Further analysis of adult looking time and puppy looking time did not support the prediction, and there

was no effect of previous experience with puppies either. It should be noted that more owners than expected (i.e., half of them) reported that their dogs had some experience with neonate puppies, which are younger than one month old and are normally kept in breeders. Thus, we need to be cautious about the validity of data of previous experience, as some owners may have reported their dog's previous experience with puppies in general rather than neonate puppies specifically. We also found evidence of a side bias; dogs looked at an image for longer presented on the right. This bias could be due to the fact that the doors were on the right side of the experimental room (even though they were visually occluded).

As always, negative results are difficult to interpret because the reasons remain ambiguous, and a multitude of factors could contribute to such an absence of evidence. First, the two dog images were presented quite far apart (approximately 2 m) for the purpose of later video coding. This may have made it difficult for dogs to notice both images. Indeed, in 68.7% of the valid trials, dogs looked at only one of the two images. Second, the dogs might not have recognized the images, especially the small puppy stimuli in the natural size condition, because of the distance between subjects and the images and given dogs' rather poor visual acuity (Miller, 1995). Third, in many trials, the dog's head was directed towards the space between the screens but not at either image, thus, it was not coded as looking at any image because we defined the looking behavior mainly by the head direction towards a screen. However, dogs may have looked at one of the stimuli without clear head orientation, which may have underestimated their actual looking times. We considered gaze direction whenever possible, but gaze direction was not always clear from the video. Although the same setting worked with other stimuli in a study by Faragó et al. (2010), it is possible that these may have caused negative results. In Experiment 2, we conducted the same experiment using eye-tracking methodology to deal with potential issues of stimulus presentation and coding of dogs' looking behavior.

Experiment 2

We tested different dogs in the same manner but with eye-tracking. Using this methodology allowed the presentation of the visual stimuli to be closer to one another and at a closer distance to the dogs, which enabled us to measure dogs' visual attention more precisely. In an exploratory analysis, we also examined the dogs' pupil size response as a proxy of their arousal levels. Pupil size can be influenced by many factors, but pupil dilation is commonly regarded as an indicator of emotional arousal (Bradley et al., 2008). Indeed, studies in dogs showed that they exhibited dilated pupils when presented with pictures of angry faces compared to happy faces (Karl et al., 2020; Somppi et al., 2017). Dogs also exhibit larger pupils following expectancy violations concerning physical regularities (Völter et al., 2023; Völter & Huber, 2021a) and their pupil size is more variable when presented with animacy cues (Völter & Huber, 2022).

Methods

Subjects

We tested 15 adult pet dogs. One male dog was excluded because of his reaction to the vocalization as described later, and the final sample size was 14 (7 females, $M = 61.6$ months, see Table A2 for details); none took part in Experiment 1. The data was collected from 14th March to 29th July 2022. The dogs had previously been trained for and had successfully participated in other eye-tracking experiments (Völter & Huber, 2021b, 2021a, 2022).

Experimental set-up and Procedure

Identical to Experiment 1, each dog experienced two sessions (i.e., same and natural size conditions) but on different days, and one session consisted of four trials. The session order was counterbalanced between subjects. The EyeLink1000 eye-tracking system (SR Research, Canada) was used to record the dogs' gaze location. Visual stimuli were presented on a 24-inch

LCD monitor (resolution: 1024 × 768; refresh rate: 60 Hz) at a distance of 70 cm from the dogs' eyes. The auditory stimuli were played back from an elongated loudspeaker (Dell AC511 Soundbar) mounted below the screen.

The dogs were already trained to put their head on a chinrest to minimize their head movement during eye-tracking. Since the dogs had never experienced any sound presentation during an eye-tracking experiment, familiarization was needed. We started by playing classical music at a low volume, then bird sounds and dog vocalizations (e.g., yawning or barking), which differed from the ones presented during the experiment. One dog was excluded at this point because he reacted to the dog vocalization so much that eye tracking was impossible. Before each session, dogs completed a 5-point calibration with animated stimuli and at least one validation procedure with an average deviation between calibration and validation of less than 1° of visual angle. Then, each trial started with presenting fixation stimuli at the center of the monitor. If the dogs' eye gaze fell into the fixation area, visual and auditory stimuli were automatically presented simultaneously. The study was approved by the institutional ethics and animal welfare committee (ETK-034/02/2022) in accordance with GPS and national legislation guidelines.

Stimuli

We used almost identical dog images as in Experiment 1, but this time the visual stimuli had a naturalistic background (Figure 3). Eye-tracking studies in dogs with static images normally employed a shorter presentation duration than 10 s (Park et al., 2022). Therefore, the presentation duration was changed to 8 s instead of 10 s in order to maintain the dogs' concentration until the end of the eye-tracking session. We used the same auditory stimuli as in Experiment 1.

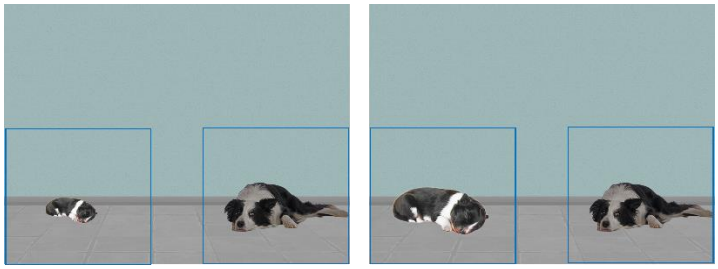

Figure 3 Examples of visual stimuli and regions of interest (*Left*: natural size condition, *Right*: same size condition)

Analysis

Two regions of interest (ROI, 430 × 400) were drawn on the dog images prior to data collection (Figure 3). Gazing was categorized as a fixation if a gaze point did not change for longer than 75 ms. The total looking time to each ROI and the image the subject first fixated (*the first look images*) were recorded. Looking time was analyzed as the adult looking proportion as in Experiment 1. The following factors were included as predictor variables: size condition (natural size/same size), vocalization (adult dog/puppy), the interaction between these two terms, the side of the adult image, the session number, the trial number, subject age, and sex. The interaction was removed if it was not significant. Subject ID was included as a random intercept, and size condition, vocalization, the side of the adult image, session number, and trial number were included as random slopes. We also analyzed adult looking time and puppy looking time divided by the total presentation duration (i.e., 8 s) using the same procedure. We also analyzed the image that the dogs looked at first in a trial using the same procedure but with binomial error distribution. We conducted likelihood ratio tests to evaluate the significance of the fixed effects. These analyses were the same as in Experiment 1, except that we did not include stimuli breed type to simplify the model due to model complexity considerations and a smaller sample size.

For the pupil size analysis, we preprocessed the data (Mathôt et al., 2018) in the following way: we inspected the data of each run in order to detect artefacts. We excluded data

100 ms prior and following blink events (detected by the Eyelink software) and applied a linear interpolation. Subsequently, we applied a subtractive baseline correction. We used the first 200 ms of the trial as the baseline period. Finally, we downsampled the data to 10 Hz to mitigate potential autocorrelation issues. To analyze the time course of pupil size, we fitted a generalized additive mixed model (GAMM) with a Gaussian error structure (following Sóskuthy, 2017; van Rij et al., 2019; and in line with our previous pupil size analyses, see Völter & Huber, 2021a, 2021b, 2022, 2023). We analyzed the remaining trial period (7.8 s) starting at the end of the baseline period. We fitted the GAMM in R using the functions 'bam' of package 'mgcv' (Wood, 2011) and package 'itsadug' (van Rij et al., 2020) to visualize the results. We used the smoothing parameter selection method 'ML'. We included the size condition, vocalization and their interaction as linear terms, the non-linear regression lines for time and for the two levels of size condition and vocalization over time (upper limit of the number of basis functions set to 30), and the non-linear interaction between X and Y gaze coordinates (because gaze position can affect pupil size Mathôt et al., 2018; van Rij et al., 2019). We also included random factor smooths for each subject and for each individual time series trajectory (i.e., for each subject and test trial) to improve the model fit and account for autocorrelation (Sóskuthy, 2017; van Rij et al., 2019).

We evaluated the model by visually inspecting correlations between the residuals and lagged residuals, the QQ-plot of residuals and the residuals against the fitted values (using the functions 'gam.check' of package 'mgcv' and 'acf' of package 'stats'). The residuals seemed to be approximately normally distributed, and there was no obvious pattern in the residuals plotted against the fitted values. However, we found some evidence for autocorrelation at lag 1 (0.675); therefore, we conducted a secondary analysis based on aggregated data (see below). To draw inferences about the effect of the predictor variables on pupil size, we first compared the full model to a reduced model excluding both the parametric and smooth terms of size condition and

vocalization using a Chi-square test of ML scores (using the function compareML of R package 'itsadug') (van Rij et al., 2020). Based on a significant full-null model comparison, we inspected the model summary and the estimates of the differences between the conditions (using the function plot_diff of R package 'itsadug') (van Rij et al., 2020). To analyze the mean baseline-corrected pupil size, we fitted a linear mixed model (using R package 'lme4'). We included size condition, vocalization, their interaction, trial number as predictor variables, subject ID as a random intercept, and all possible random slope components. The data, R code and a video are available at Mendeley Data (doi: 10.17632/6skk6w8tnb.1).

Results

We examined whether dogs' looking behavior (the adult looking proportion and the first look) was influenced by vocalization types. Figure 4 shows the adult looking proportion. The GLMM for the adult looking proportion revealed no significant interaction effect between size condition and vocalization (beta GLMM, LR test: χ^2 (1) = 0.02, p = 0.88). After removing the interaction, we found no effect of vocalization (χ^2 (1) = 0.03, p = 0.86, Table A7 and A8) but found a significant effect of size condition (χ^2 (1) = 6.40, p = 0.01). Namely, the proportion of looking time directed at adult images compared to puppy images was again longer in the natural size condition than in the same size condition, but it was not affected by the vocalization type. A binomial GLMM of the first look images also provided no evidence for a significant interaction between the size condition and vocalization type (LR test: χ^2 (1) = 0.05, p = 0.82). After removing the interaction, we found a significant effect of size condition (χ^2 (1) = 5.66, p = 0.02, Figure 4, Table A7 and A8), namely, the likelihood of dogs looking at the adult image first was higher when

it was larger than the puppy image (natural size condition) than when they were the same size (same size condition). There was no effect of vocalization, though (χ^2 (1) = 0.07, p = 0.79). When analyzing adult looking time and puppy looking time divided by the total presentation duration, the main effect of vocalization was not significant in either index (beta GLMM, LR test: adult looking time: $\chi^2(1)$ =085, p = 0.36, puppy looking time: $\chi^2(1)$ =0.42, p =0.51, Table A9 for details). Thus, the results in Experiment 2 were rather similar to Experiment 1: the looking time and the first look were influenced by size condition but not vocalization. Unlike in Experiment 1, in Experiment 2 we did not find a significant effect of the adult side in any analysis. We did not analyze previous experience with puppies in Experiment 2 because only one subject had previous experience. In 43.4% of all trials, dogs looked at both ROIs.

Based on visual inspection of the data, the overall looking time was seemingly longer when the vocalization was from a puppy compared to from an adult, especially in the same size condition (Figure A2). Therefore, in an exploratory analysis, we also analyzed "the overall looking time," which is the sum of the looking time of the adult dog and the puppy stimuli. We used a non-parametric analysis (Wilcoxon signed rank test) to compare the overall looking time between the two vocalization types for each size condition since the looking time was not normally distributed. In the same size condition, the overall looking time was significantly longer when the vocalization was from a puppy (mean ± SD: 5270 ± 2342 ms), than from an adult (3978 ± 2759 ms) (n=14, T^+ = 11, p<0.01), while in the natural size condition, the vocalization type did not affect the overall looking time (puppy: 4158 ± 2414 ms , adult: 3946 ± 1372 ms, n = 14, T^+ = 44, p = 0.63).

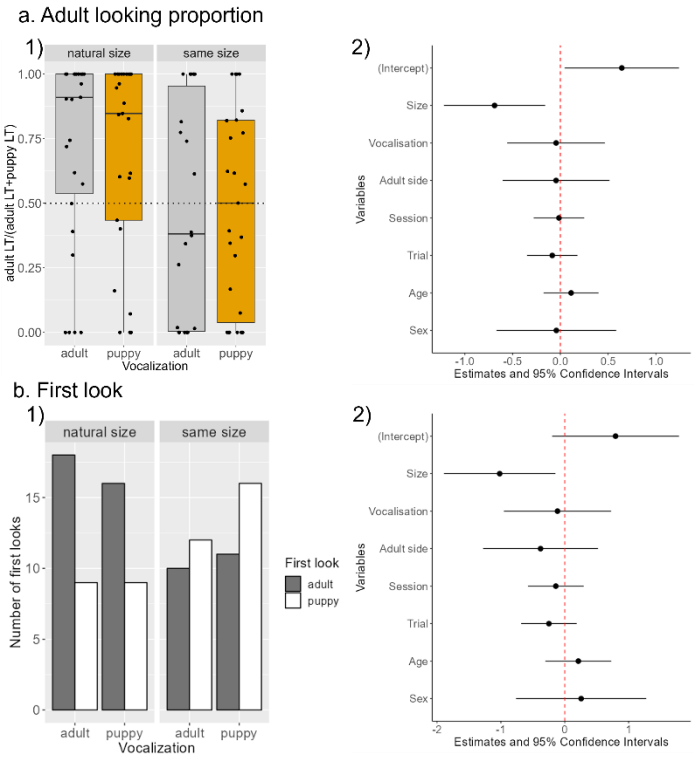

Figure 4 The results of looking behavior in Experiment 2

a-1) Boxplot showing the proportion of looking at the adult dogs (adult looking time/ (adult looking time + puppy looking time). The dots represent the data point for each trial. Scores greater than 0.5 in the adult looking proportion indicate that a subject looked at an adult dog for a longer time than at a puppy. a-2) Estimate of fixed effects and 95% confidence interval of the GLMM modelling the adult looking proportion.

b-1) The total number of first looks from all subjects. b-2) Estimate of fixed effects and 95% confidence interval of the GLMM modelling the first looks.

The increase of overall looking time when hearing a puppy vocalization may indicate that dogs were aroused by the vocalization. Investigating a similar question, we also checked pupillometry as another indicator of dogs' arousal. We analyzed the preprocessed, baseline-corrected pupil size data by fitting a GAMM. The full model (GAMM01), including size condition, vocalization, and their interaction as parametric terms and the non-parametric regression lines for the two levels of size condition and vocalization over time (as well as control

terms and random effects) fitted the data significantly better than the null model that lacked these test predictors (Chi-square test of ML scores: $\chi^2(11) = 38.91$, $p < 0.001$; GAMM01 had a lower AIC: ΔAIC 130.73). However, the interaction between size condition and vocalization failed to reach significance ($t = -1.60$, $p = 0.111$). We then fitted the model without this interaction term (GAMM02; Chi-square test of ML scores: $\chi^2(10) = 37.65$, $p < 0.001$; GAMM02 had a lower AIC: ΔAIC 131.07). The pupil size was significantly larger when the dogs were presented with a puppy than with an adult vocalization ($t = 2.01$, $p = 0.044$, Figure 5a). The size condition had no significant effect on pupil size ($t = 0.34$, $p = 0.731$). The difference curve confirmed that pupil size was significantly larger in puppy than in adult vocalization trials in the time window between 2689 and 3856 ms (Figure 5b).

In a secondary analysis, we first aggregated the data per individual and trial (mean values) and then fitted an LMM with size condition, vocalization and trial number as predictors. We found a significant interaction between size condition and vocalization ($\chi^2(1) = 5.35$, $p = 0.021$; for the model estimates, see Table A10). Post-hoc pairwise comparisons revealed that the dogs' pupils were, on average, significantly larger when presented with the puppy vocalization than the adult vocalization but only in the natural size condition (t (38.36) = 2.48, $p = 0.018$) and not in the same size condition (t (38.02) = -0.35, $p = 0.731$). The reason for the differences in the results between the first analysis (no effect of size condition) and this one (a significant interaction effect between size condition and vocalization) is not entirely clear but might be related to the fact that the GAMM allows for taking nonlinear effects over time (alongside parametric effects) into account and for including control predictors such as the current gaze coordinates (to account for the pupil foreshortening effect).

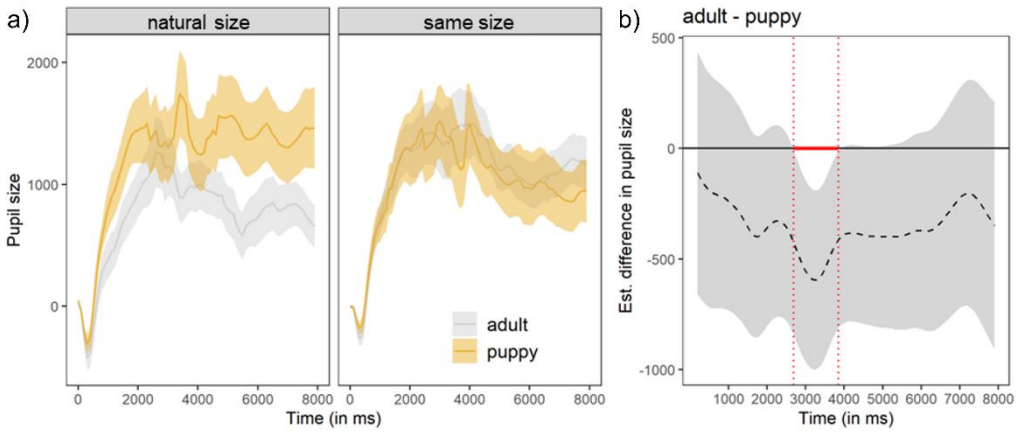

Figure 5 Pupil size data over time in Experiment 2

a) Time series plot showing dogs' pupil size (in arbitrary units and baseline corrected). The orange and light grey lines show the mean pupil size (± se, shaded area around the line) in the puppy and adult vocalization conditions. The left panel shows the data of the natural size condition; the right panel shows data of the same size condition.

b) Difference curve derived from GAMM02. The dashed line shows the estimated difference between the adult and puppy vocalization conditions; the shaded area shows the pointwise 95% confidence interval. The period in which the conditions significantly differ is highlighted in red (delimited by the dotted line).

Discussion

The results of Experiment 2 are similar to the results of Experiment 1. The vocalization type did not affect dogs' looking time nor the location of first looks. The only pattern we found is that the dogs looked at adult dogs for longer and quicker when an adult dog was presented at a larger size than a puppy, indicating that dogs paid more attention to physically more salient stimuli. Therefore, there is no evidence that dogs spontaneously match static images and vocalizations based on conspecific age class. We also found that the overall looking time increased (in the same size condition) and pupils were dilated (in the natural size condition) when dogs heard the vocalizations of puppies compared to adult dogs. These results suggest that dogs were more aroused and attentive when they heard puppy vocalizations.

582

583 ## General Discussion

584 This study investigated whether dogs can spontaneously match auditory and visual cues
585 regarding the age class of conspecifics in a cross-modal matching task with two methods. In
586 Experiment 1, based on the methods of Faragó et al. (2010), the visual stimuli were projected on
587 a wall, and the dogs' looking behavior was determined by their head direction. In Experiment 2,
588 we applied eye tracking and analyzed dogs' looking direction in more detail using the same stimuli
589 as in Experiment 1. The results were consistent across both experiments. In both experiments,
590 dogs did not differentiate between cross-modally matched and mismatched stimuli by their
591 looking behavior. Thus, we did not find evidence of dogs having cross-modal representations of
592 conspecific age classes. The only consistent result we found was that dogs looked at an adult dog
593 for longer and quicker when it was presented larger than a puppy (i.e., in the natural size
594 condition).

595 Our results provide no support that dogs recognize conspecific age classes from static
596 images or vocalizations and integrate them. There are the following three possibilities: a) dogs
597 cannot use visual cues and/or b) dogs cannot use auditory cues, or c) dogs can use both visual and
598 auditory cues but this experimental paradigm did not allow them to exhibit it.

599 First, considering the possibility that dogs cannot use visual cues, a previous eye-
600 tracking study, in which pairs of human and dog faces were presented accompanied by either a
601 dog's bark or human speech, has shown that dogs can match images and vocalizations of
602 conspecifics versus humans (Gergely et al., 2019). Another study also showed that dogs can
603 visually differentiate conspecifics of various breeds from other species, despite the great
604 phenotypic variability among dogs (Autier-Dérian et al., 2013). Based on these studies, it is
605 unlikely that our subjects did not recognize at least the adult dogs as dogs in the presented images.

However, it is possible that identifying age class from still images of other dogs is challenging for dogs because of the considerable variation in appearance among breeds due to artificial breeding. Such variability including body size across breeds may make age class categorization purely based on their appearance difficult. Moreover, paedomorphic visual features (e.g., dolichocephalic head) remain even in mature dogs, especially for some breeds. Thus, static conspecific visual appearance may not be a primary cue of age for dogs; instead, they may need other cues, such as behavior or smell, to detect whether a conspecific is a puppy or an adult dog.

Another possibility is that the visual stimuli we used were insufficient. For example, video stimuli could be more informative to the dogs and also may attract attention more effectively to both stimuli. Indeed, in a previous study with rhesus macaques, where a cross-modal representation of age-related body size was found, video stimuli were used (Ghazanfar et al., 2007). However, in the present study, we opted for static images because we were not able to acquire or produce well-controlled videos for both adult dogs and puppies (considering factors like breed, movement, situation, posture, etc.). Nevertheless, it should be noted that a growing number of studies have shown that dogs can recognize still images presented on a screen, evidenced by their performance in cross-modal matching studies (e.g., Adachi et al., 2007; Faragó, Pongrácz, Miklósi, et al., 2010; Gergely et al., 2019), categorization tasks (e.g., Huber et al., 2013; Range et al., 2008), and expectancy violation studies (e.g., Völter & Huber, 2021a, 2021b, 2022, 2023).

It is also possible that dogs cannot use auditory cues to recognize age class. In previous studies by Faragó, Pongrácz, Miklósi, et al., (2010) and Taylor et al. (2011), dogs successfully matched a dog image and a growl with respect to body size. In Experiment 1, we used the same methodology as Faragó and colleagues, and the body size difference was an easily recognizable feature of our two age classes in the natural size condition. Therefore, it may seem unexpected

that dogs did not match age class even in our natural size condition. There could be multiple reasons for the discrepancy between our current results and these previous papers; among them is the different acoustic stimuli used. In the previous studies, growls were used in contrast to the whining stimuli used in the current study. The most important acoustic feature that encodes body size information in dogs is formant dispersion, which is prominent in broad-band growl but not in narrow-band whining (Riede & Fitch, 1999; Taylor et al., 2008). Even if dogs' body size affects the acoustic features of whining (Sibiryakova et al., 2021), it may be rather subtle and insufficient for conspecifics to estimate the caller's body size. Acquiring body size information may be more important when dogs produce growls, which are mainly produced during agonistic interactions as a threatening signal（Faragó et al., 2010) compared to when dogs emit whine upon separation from the owner or their mother. Particularly, agonistic growls are expected to be more salient in this context, possibly alerting the listener and providing information about the level of threat, whereas whines may convey the distress of the caller and elicit empathetic responses through emotional contagion (Quervel-Chaumette et al., 2016), thus potentially less crucial for them to convey size information.

In the natural size condition, dogs showed a strong looking bias for adult dogs. Looking bias for larger dog images has also been reported in another cross-modal matching study in dogs (Bálint et al., 2013). They modified Faragó et al.'s (2010) paradigm and conducted a similar cross-modal body size matching study using playful and agonistic growls as auditory stimuli. They reported that dogs looked at larger dog images more than at smaller ones after hearing playful growls. The authors suggested that playful growls might contain exaggerated size information, which could explain their results (Faragó, Pongrácz, Range, et al., 2010). However, we used whines in the present study, so dogs' looking bias for adult dogs in the natural condition would probably have occurred simply due to their physical saliency as we discuss below. Nevertheless,

it should be noted that we found that dogs showed looking time and pupil dilation differences when hearing puppy vocalizations and adult vocalizations. This suggests that dogs at least differentiated the two types of whining.

Alternatively, dogs may be able to use visual and auditory cues of age class, but the experimental setting may have not allowed dogs to demonstrate their ability. In other words, it is possible that even though dogs can recognize age categories from visual and auditory cues, it may have not systematically affected their looking behavior. For example, the robust size difference between adult dogs and puppies in the natural size condition may have overridden the congruency effect if any. In both experiments, we consistently observed that dogs' visual attention tended to be biased toward the adult image in the natural size condition. In a prior study by Faragó et al. (2010), the size difference between a pair of dog images was 30%. In contrast, the adult dog images in the present study were much larger than puppy images (i.e., three times bigger in height) in the natural size condition to make the relative size as realistic as possible. Therefore, in our study, body size could be the predominant factor influencing dogs' visual attention. Adult dogs may have simply been physically much more salient, and they may have attracted dogs' attention. Hence, while dogs might possess the ability to discern age classes using visual and auditory cues in their daily life, this ability might not have been fully reflected in their looking behavior within our experimental setup. Instead, the result may reflect the complex nature of cross-modal studies. Although most cross-modal studies with paired visual stimuli have reported looking bias for congruent stimuli, there is a study that reported a bias for incongruent stimuli (Jardat et al., 2022), and some have reported both a looking bias for congruent and incongruent stimuli within a study depending on the previous experience of the subject (Ratcliffe et al., 2014) or experimental stimuli (Zangenehpour et al., 2009). Thus, we found no looking bias for congruent stimuli possibly because a looking bias for congruent stimuli was diminished or overwritten by avoiding congruent

678 stimuli (Ratcliffe et al., 2014; Zangenehpour et al., 2009) or a looking bias for incongruent stimuli
679 due to expectancy violation (Jardat et al., 2022).

680 We also asked the owners about their dogs' previous experience with puppies. We used
681 very young puppies as stimuli because we wanted to use stimuli of clearly different age classes.
682 However, dogs rarely have a chance to meet such young puppies. In Experiment 2, indeed, all the
683 subjects except one had no previous experience with neonatal puppies. Previous research on
684 human infants has shown that experience can make a difference in age-class categorizations:
685 Bahrick et al. (2008) showed that 7-month-olds can match the faces and voices of adults and
686 children, and previous experience with children facilitated this matching. However, in Experiment
687 1, dogs' previous experience with puppies did not affect the congruency effect. Also, no effect of
688 previous experience was reported on dogs' attentional response to puppy vocalizations in a
689 previous study (Lehoczki et al., 2019). Therefore, lack of previous experience may not fully
690 explain our results.

691 From an evolutionary point of view, one possibility for our negative results concerning
692 age-class recognition in dogs might be related to their domestication. In wolves, all pack members,
693 not just the mother, participate in pup-rearing, while in dogs, pups are solely reared by the mother
694 with varying degrees of help from human caretakers (Lord et al., 2013; Marshall-Pescini et al.,
695 2017). Due to this domestication-related change and direct human interferences in breeding, dogs
696 might have reduced sensitivity toward puppy stimuli. As a result, conspecific age information
697 may be less important for dogs than their wild counterpart. Nevertheless, dogs did not entirely
698 lose their responsivity toward puppy stimuli throughout their domestication. We found that dogs
699 looked at conspecific images longer (in the same size condition) and dilated their pupils to a
700 greater extent (in the natural size condition) after hearing puppy vocalizations in Experiment 2.
701 These two changes likely indicate that dogs were aroused by puppy vocalizations. In humans,

pupil dilation is also reported in response to stimuli depicting infant distress or discomfort (Yrttiaho et al., 2017). Previous studies have also shown that dogs show an emotional response to separation calls of puppies (Lehoczki et al., 2019, 2020; Root-Gutteridge et al., 2021). In the current study, we provide evidence that dogs' arousal increased selectively in response to puppy whining. Infant distress vocalizations in mammals have specific acoustic features which have behavioral and physiological influence on caregivers (Lingle et al., 2012). Our results indicate that the specific acoustic features of puppy whining (but not whining in general) have psychophysiological effects in dogs.

The present study tested dogs' cross-modal representation of age class, which has only been tested, to our knowledge, in limited species such as rhesus macaques (Ghazanfar et al., 2007), Australia sea lions (Charrier et al., 2022), and horses (not conspecific but human age class, Jardat et al., 2022) besides humans (e.g., Bahrick et al., 2008). Our results provide no evidence of cross-modal age representations in dogs. Apart from methodological limitations, one potential explanation for dogs' performance is that they may need other cues to recognize the conspecific age class. Additionally, our study highlighted the possibility that puppy whining specifically induces elevated arousal in dogs. This result is not only consistent with but adds to previous results by showing that this arousal effect is selectively induced by puppy whining as compared to adult whining.

CRediT statement

Yuri Kawaguchi: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization, Project administration, Funding acquisition. **Zsófia Virányi**: Conceptualization, Methodology, Writing - Review & Editing. **Tamás Faragó**: Conceptualization, Methodology, Writing - Review & Editing. **Ludwig Huber**: Conceptualization, Methodology, Resources, Writing - Review & Editing, Supervision. **Christoph J. Völter**: Conceptualization, Methodology, Software, Formal analysis, Writing - Review & Editing, Visualization.

730 References

731 Adachi, I., Kuwahata, H., & Fujita, K. (2007). Dogs recall their owner's face upon hearing the
732 owner's voice. *Animal Cognition*, *10*(1), 17–21. https://doi.org/10.1007/s10071-006-
733 0025-8

734 Adachi, I., Kuwahata, H., Fujita, K., Tomonaga, M., & Matsuzawa, T. (2009). Plasticity of ability
735 to form cross-modal representations in infant Japanese macaques. *Developmental Science*,
736 *12*(3), 446–452. https://doi.org/10.1111/j.1467-7687.2008.00780.x

737 Autier-Dérian, D., Deputte, B. L., Chalvet-Monfray, K., Coulon, M., & Mounier, L. (2013). Visual
738 discrimination of species in dogs (*Canis familiaris*). *Animal Cognition*, *16*(4), 637–651.
739 https://doi.org/10.1007/s10071-013-0600-8

740 Bahrick, L. E., Netto, D., & Hernandez-Keif, M. (2008). Intermodal perception of adult and child
741 faces and voices by infants. *Child Development*, *69*(5), 1263–1275.
742 https://doi.org/10.1111/j.1467-8624.1998.tb06210.x

743 Bálint, A., Faragó, T., Dóka, A., Miklósi, Á., & Pongrácz, P. (2013). 'Beware, I am big and non-
744 dangerous!' – Playfully growling dogs are perceived larger than their actual size by their
745 canine audience. *Applied Animal Behaviour Science*, *148*(1–2), 128–137.
746 https://doi.org/10.1016/j.applanim.2013.07.013

747 Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for

confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Berry, D. S., & McArthur, L. Z. (1986). Perceiving character in faces: The impact of age-related craniofacial changes on social perception. *Psychological Bulletin*, *100*(1), 3–18. https://doi.org/10.1037/0033-2909.100.1.3

Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology, 45*(4), 602-607

Brooks, M., E., Kristensen, K., Benthem, K., J. ,van, Magnusson, A., Berg, C., W., Nielsen, A., Skaug, H., J., Mächler, M., & Bolker, B., M. (2017). GlmmTMB Balances speed and flexibility among packages for zero-inflated Generalized Linear Mixed Modeling. *The R Journal*, *9*(2), 378. https://doi.org/10.32614/RJ-2017-066

Charrier, I., Pitcher, B. J., & Harcourt, R. (2022). Mother–pup recognition mechanisms in Australia sea lion (*Neophoca cinerea*) using uni- and multi-modal approaches. *Animal Cognition*, *25*, 1019–1028.

Cohen, J. A., & Fox, M. W. (1976). Vocalizations in wild canids and possible effects of domestication. *Behavioural Processes, 1*(1), 77–92. https://doi.org/10.1016/0376-

6357(76)90008-5

Faragó, T., Pongrácz, P., Miklósi, Á., Huber, L., Virányi, Z., & Range, F. (2010). Dogs' expectation about signalers' body size by virtue of their growls. *PLoS ONE*, *5*(12), e15175. https://doi.org/10.1371/journal.pone.0015175

Faragó, T., Pongrácz, P., Range, F., Virányi, Z., & Miklósi, Á. (2010). 'The bone is mine': Affective and referential aspects of dog growls. *Animal Behaviour*, *79*(4), 917–925. https://doi.org/10.1016/j.anbehav.2010.01.005

Gergely, A., Petró, E., Oláh, K., & Topál, J. (2019). Auditory–visual matching of conspecifics and non-conspecifics by dogs and human infants. *Animals*, *9*(1), 17. https://doi.org/10.3390/ani9010017

Ghazanfar, A. A., Turesson, H. K., Maier, J. X., van Dinther, R., Patterson, R. D., & Logothetis, N. K. (2007). Vocal-tract resonances as indexical cues in rhesus monkeys. *Current Biology*, *17*(5), 425–430. https://doi.org/10.1016/j.cub.2007.01.029

Huber, L., Racca, A., Scaf, B., Virányi, Z., & Range, F. (2013). Discrimination of familiar human faces in dogs (*Canis familiaris*). *Learning and Motivation*, *44*(4), 258–269. https://doi.org/10.1016/j.lmot.2013.04.005

Jardat, P., Ringhofer, M., Yamamoto, S., Gouyet, C., Degrande, R., Parias, C., Reigner, F., Calandreau, L., & Lansade, L. (2022). Horses form cross-modal representations of adults

784 and children. *Animal Cognition*. https://doi.org/10.1007/s10071-022-01667-9

785 Karl, S., Boch, M., Zamansky, A., van der Linden, D., Wagner, I. C., Völter, C. J., Lamm, C., &

786 Huber, L. (2020). Exploring the dog–human relationship by combining fMRI, eye-

787 tracking and behavioural measures. *Scientific Reports*, *10*(1).

788 https://doi.org/10.1038/s41598-020-79247-5

789 Lehoczki, F., Szamosvölgyi, Z., Miklósi, Á., & Faragó, T. (2019). Dogs' sensitivity to strange pup

790 separation calls: Pitch instability increases attention regardless of sex and experience.

791 *Animal Behaviour*, *153*, 115–129. https://doi.org/10.1016/j.anbehav.2019.05.010

792 Lehoczki, F., Szenczi, P., Bánszegi, O., Jakobsen, K., & Faragó, T. (2020). Cross-species effect

793 of separation calls: Family dogs' reactions to pup, baby, kitten and artificial sounds.

794 *Animal Behaviour, 168*, 169-185.

795 Lingle, S., Wyman, M. T., Kotrba, R., Teichroeb, L. J., & Romanow, C. A. (2012). What makes a

796 cry a cry? A review of infant distress vocalizations. *Current Zoology*, *58*(5), 698–726.

797 https://doi.org/10.1093/czoolo/58.5.698

798 Lord, K., Feinstein, M., Smith, B., & Coppinger, R. (2013). Variation in reproductive traits of

799 members of the genus Canis with special attention to the domestic dog (*Canis familiaris*).

800 *Behavioural Processes*, *92*, 131–142. https://doi.org/10.1016/j.beproc.2012.10.009

801 Marshall-Pescini, S., Cafazzo, S., Virányi, Z., & Range, F. (2017). Integrating social ecology in

802 explanations of wolf–dog behavioral differences. *Current Opinion in Behavioral Sciences*,
803 *16*, 80–86. https://doi.org/10.1016/j.cobeha.2017.05.002
804 Marx, A., Lenkei, R., Pérez Fraga, P., Bakos, V., Kubinyi, E., & Faragó, T. (2021). Occurrences
805 of non-linear phenomena and vocal harshness in dog whines as indicators of stress and
806 ageing. *Scientific Reports*, *11*(1), 4468. https://doi.org/10.1038/s41598-021-83614-1
807 Mathôt, S., Fabius, J., Van Heusden, E., & Van der Stigchel, S. (2018). Safe and sensible
808 preprocessing and baseline correction of pupil-size data. *Behavior Research Methods*,
809 *50*(1), 94–106. https://doi.org/10.3758/s13428-017-1007-2
810 Miller, P. E. (1995). Vision in dogs. *Journal of the American Veterinary Medical Association*.
811 *207*(12), 1623–1634.
812 Mongillo, P., Eatherington, C., Lõoke, M., & Marinelli, L. (2021). I know a dog when I see one:
813 Dogs (*Canis familiaris*) recognize dogs from videos. *Animal Cognition, 24,* 969-979.
814 https://doi.org/10.1007/s10071-021-01470-y
815 Park, S. Y., Holmqvist, K., Niehorster, D. C., Huber, L., & Virányi, Z. (2022). How to improve
816 data quality in dog eye tracking. *Behavior Research Methods, 55*(4), 1513-1536.
817 https://doi.org/10.3758/s13428-022-01788-6
818 Proops, L., McComb, K., & Reby, D. (2009). Cross-modal individual recognition in domestic
819 horses (*Equus caballus*). *Proceedings of the National Academy of Sciences*, *106*, 947–

951.

Quervel-Chaumette, M., Faerber, V., Faragó, T., Marshall-Pescini, S., & Range, F. (2016). Investigating empathy-like responding to conspecifics' distress in pet dogs. *PLoS ONE*, *11*(4), e0152920. https://doi.org/10.1371/journal.pone.0152920

Range, F., Aust, U., Steurer, M., & Huber, L. (2008). Visual categorization of natural stimuli by domestic dogs. *Animal Cognition*, *11*(2), 339–347. https://doi.org/10.1007/s10071-007-0123-2

Ratcliffe, V. F., McComb, K., & Reby, D. (2014). Cross-modal discrimination of human gender by domestic dogs. *Animal Behaviour*, *91*, 127–135. https://doi.org/10.1016/j.anbehav.2014.03.009

R Core Team. (2023). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing (Version 3.5.1). http://www.r-project.org/

Riede, T., & Fitch, T. (1999). Vocal tract length and acoustics of vocalization in the domestic dog (*Canis familiaris*). *The Journal of Experimental Biology*, *202*, 2859–2867.

Root-Gutteridge, H., Ratcliffe, V. F., Neumann, J., Timarchi, L., Yeung, C., Korzeniowska, A. T., Mathevon, N., & Reby, D. (2021). Effect of pitch range on dogs' response to conspecific vs. Heterospecific distress cries. *Scientific Reports*, *11*(1), 19723.

https://doi.org/10.1038/s41598-021-98967-w

Sibiryakova, O. V., Volodin, I. A., & Volodina, E. V. (2021). Polyphony of domestic dog whines and vocal cues to body size. *Current Zoology*, *67*(2), 165–176. https://doi.org/10.1093/cz/zoaa042

Sliwa, J., Duhamel, J.-R., Pascalis, O., & Wirth, S. (2011). Spontaneous voice-face identity matching by rhesus monkeys for familiar conspecifics and humans. *Proceedings of the National Academy of Sciences*, *108*(4), 1735–1740. https://doi.org/10.1073/pnas.1008169108

Somppi, S., Törnqvist, H., Topál, J., Koskela, A., Hänninen, L., Krause, C. M., & Vainio, O. (2017). Nasal oxytocin treatment biases dogs' visual attention and emotional response toward positive human facial expressions. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01854

Sóskuthy, M. (2017). *Generalised additive mixed models for dynamic analysis in linguistics: A practical introduction* (arXiv:1703.05339). arXiv. http://arxiv.org/abs/1703.05339

Takagi, S., Arahori, M., Chijiiwa, H., Saito, A., Kuroshima, H., & Fujita, K. (2019). Cats match voice and face: Cross-modal representation of humans in cats (*Felis catus*). *Animal Cognition*, *22*(5), 901–906. https://doi.org/10.1007/s10071-019-01265-2

Taylor, A. M., Reby, D., & McComb, K. (2008). Human listeners attend to size information in

domestic dog growls. *The Journal of the Acoustical Society of America*, *123*(5), 2903–2909. https://doi.org/10.1121/1.2896962

Taylor, A. M., Reby, D., & McComb, K. (2011). Cross modal perception of body size in domestic dogs (*Canis familiaris*). *PLoS ONE*, *6*(2), 6.

van Rij, J., Hendriks, P., van Rijn, H., Baayen, R. H., & Wood, S. N. (2019). Analyzing the time course of pupillometric data. *Trends in Hearing*, *23*, 233121651983248. https://doi.org/10.1177/2331216519832483

van Rij, J., Wieling, M., Baayen, R. H., & van Rijn, H. (2020). *itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs*.

Völter, C. J., & Huber, L. (2021a). Dogs' looking times and pupil dilation response reveal expectations about contact causality. *Biology Letters*, *17*(12), 20210465. https://doi.org/10.1098/rsbl.2021.0465

Völter, C. J., & Huber, L. (2021b). Expectancy violations about physical properties of animated objects in dogs. *Proceedings of the Annual Meeting of the Cognitive Science S*, *43*. https://doi.org/10.31234/osf.io/3pr9z

Völter, C. J., & Huber, L. (2022). Pupil size changes reveal dogs' sensitivity to motion cues. *IScience*, 104801. https://doi.org/10.1016/j.isci.2022.104801

Völter, C. J., Tomašić, A., Nipperdey, L., & Huber, L. (2023). Dogs' expectations about occlusion

874 events: From expectancy violation to exploration. *Proceedings of the Royal Society B:*
875 *Biological Sciences*, 290(2003), 20230696. https://doi.org/10.1098/rspb.2023.0696
876 Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation
877 of semiparametric generalized linear models: Estimation of Semiparametric Generalized
878 Linear Models. *Journal of the Royal Statistical Society: Series B (Statistical*
879 *Methodology)*, *73*(1), 3–36. https://doi.org/10.1111/j.1467-9868.2010.00749.x
880 Yeon, S. C. (2007). The vocal communication of canines. *Journal of Veterinary Behavior*, 2(4), 141–
881 144. https://doi.org/10.1016/j.jveb.2007.07.006
882 Yong, M. H., & Ruffman, T. (2014). Emotional contagion: Dogs and humans show a similar
883 physiological response to human infant crying. *Behavioural Processes*, *108*, 155–165.
884 https://doi.org/10.1016/j.beproc.2014.10.006
885 Yrttiaho, S., Niehaus, D., Thomas, E., & Leppänen, J. M. (2017). Mothers' pupillary responses to
886 infant facial expressions. *Behavioral and Brain Functions*, *13*(1), 2–2.
887 https://doi.org/10.1186/s12993-017-0120-9
888 Jardat, P., Ringhofer, M., Yamamoto, S., Gouyet, C., Degrande, R., Parias, C., Reigner, F.,
889 Calandreau, L., & Lansade, L. (2022). Horses form cross-modal representations of adults
890 and children. *Animal Cognition*. https://doi.org/10.1007/s10071-022-01667-9
891 Ratcliffe, V. F., McComb, K., & Reby, D. (2014). Cross-modal discrimination of human gender

892 by domestic dogs. *Animal Behaviour*, *91*, 127–135.
893 https://doi.org/10.1016/j.anbehav.2014.03.009
894 Zangenehpour, S., Ghazanfar, A. a, Lewkowicz, D. J., & Zatorre, R. J. (2009). Heterochrony and
895 cross-species intersensory matching by infant vervet monkeys. *PloS One*, *4*(1), e4302–
896 e4302. https://doi.org/10.1371/journal.pone.0004302
897